

Licensing software as part of an EDI data package

Mark Servilla, Draft version 1.0, 10 September 2020

Introduction

The following document prescribes best practices for licensing software that is contributed as part of a “data package” submitted to the Environmental Data Initiative (EDI) data repository. Such software **should** have a direct and impactful relationship to one or more data entity components of the data package (e.g., software that creates, updates, or analyzes a data entity).

Disclaimer

This document does not assume to be a legal guide, nor does it represent legal code, but rather is only a recommendation for licensing software in the context of a “data package” as archived and published through the EDI data repository. Ultimately, software licensing and its impact on reuse is the sole responsibility of the “data package” author.

Definitions

1. **Copyright** - “Copyright is a type of intellectual property that gives its owner the exclusive right to make copies of a creative work, usually for a limited time. The creative work may be in a literary, artistic, educational, or musical form. Copyright is intended to protect the original expression of an idea in the form of a creative work, but not the idea itself.” - Wikipedia, [Copyright](#), 2020-09-6. The following sections pertain specifically to [Copyright law of the United States of America](#):
 - a. **Subject matter of copyright** - “Copyright protection subsists, in accordance with this title, in original works of authorship fixed in any tangible medium of expression, now known or later developed, from which they can be perceived, reproduced, or otherwise communicated, either directly or with the aid of a machine or device. Copyright.gov, [17 U.S. Code § 102 - Subject matter of copyright: In general](#), 2020-09-6.
 - b. **A “work”** - A work is “created” when it is fixed in a copy or phonorecord for the first time; where a work is prepared over a period of time, the portion of it that has been fixed at any particular time constitutes the work as of that time, and where the work has been prepared in different versions, each version constitutes a separate work. - Copyright.gov, [17 U.S. Code § 101 - Definitions](#), 2020-09-6.
 - c. **Literary works** - “*Literary works* are works, other than audiovisual works, expressed in words, numbers, or other verbal or numerical symbols or indicia, regardless of the nature of the material objects, such as books, periodicals, manuscripts, phonorecords, film,

tapes, disks, or cards, in which they are embodied.” - Copyright.gov, [17 U.S. Code § 101 - Definitions](#), 2020-09-6.

- d. **Computer program** - “A *computer program* is a set of statements or instructions to be used directly or indirectly in a computer in order to bring about a certain result.” - Copyright.gov, [17 U.S. Code § 101 - Definitions](#), 2020-09-6, 2020-09-6.
 - e. **Duration of copyright: Works created on or after January 1, 1978** - “Copyright in a work created on or after January 1, 1978, subsists from its creation and, except as provided by the following subsections, endures for a term consisting of the life of the author and 70 years after the author’s death.” and “In the case of a joint work prepared by two or more authors who did not work for hire, the copyright endures for a term consisting of the life of the last surviving author and 70 years after such last surviving author’s death.” Copyright.gov, [Duration of copyright: Works created on or after January 1, 1978](#), 2020-09-06.
2. **Software license** - “A software license is a legal instrument (usually by way of contract law, with or without printed material) governing the use or redistribution of software. Under United States copyright law, all software is copyright protected, in both source code and object code forms, unless that software was developed by the United States Government, in which case it cannot be copyrighted. Authors of copyrighted software can donate their software to the public domain, in which case it is also not covered by copyright and, as a result, cannot be licensed.
- A typical software license grants the licensee, typically an end-user, permission to use one or more copies of software in ways where such a use would otherwise potentially constitute copyright infringement of the software owner’s exclusive rights under copyright.” - Wikipedia, [Software License](#), 2020-09-05.
3. **Open source software** - “Open Source software is software that can be freely accessed, used, changed, and shared (in modified or unmodified form) by anyone. Open source software is made by many people, and distributed under licenses that comply with the [Open Source Definition](#).” - OpenSource.org, [What is “Open Source” software?](#), 2020-09-05
4. **Free and open licenses:**
- a. **Public Domain** - “The public domain consists of all the creative work to which no exclusive intellectual property rights apply. Those rights may have expired, been forfeited, expressly waived, or may be inapplicable.” - Wikipedia, [Public Domain](#), 2020-09-05.
 - b. **Permissive** - “A permissive software license, sometimes also called BSD-like or BSD-style license, is a free-software license with only minimal restrictions on how the software can be used, modified, and redistributed, usually including a warranty disclaimer.” - Wikipedia, [Permissive software license](#), 2020-09-05.
 - c. **Copyleft** - “Copyleft is the practice of granting the right to freely distribute and modify intellectual property with the requirement that the same rights be preserved in derivative works created from that property.” - Wikipedia, [Copyleft](#), 2020-09-06.
5. **Creative Commons** - “[Creative Commons](#) is a global nonprofit organization that enables sharing and reuse of creativity and knowledge through the provision of free legal tools. Our legal tools help those who want to encourage reuse of their works by offering them for use under generous, standardized terms; those who want to make creative uses of works; and those who want to benefit from this symbiosis.” and “Although Creative Commons is best known for its licenses, our work extends beyond just providing copyright licenses. CC offers other legal and technical tools that also facilitate sharing and discovery of creative works,

such as CC0, a public domain dedication for rights holders who wish to put their work into the public domain before the expiration of copyright, and the Public Domain Mark, a tool for marking a work that is in the worldwide public domain. Creative Commons licenses and tools were designed specifically to work with the web, which makes content that is offered under their terms easy to search for, discover, and use.” - Creative Commons, [What is Creative Commons and what do you do?](#), 2020-0906.

The following are [simple definitions](#) for Creative Commons licenses Version 4.0 (detailed information, including legal code, may be found on the Creative Commons [wiki](#)):

- a. **CC0** - “CC0 is the *no copyright reserved* option in the Creative Commons toolkit - it effectively means relinquishing all copyright and similar rights that you hold in a work and dedicating those rights to the public domain.” Creative Commons, [CC0](#), 2020-09-06.
 - b. **CC BY** - “This license lets others distribute, remix, adapt, and build upon your work, even commercially, as long as they credit you for the original creation. This is the most accommodating of licenses offered. Recommended for maximum dissemination and use of licensed materials.” Creative Commons, [CC BY](#), 2020-09-06.
 - c. **CC BY-SA** - “This license lets others remix, adapt, and build upon your work even for commercial purposes, as long as they credit you and license their new creations under the identical terms. This license is often compared to *copyleft* free and open source software licenses. All new works based on yours will carry the same license, so any derivatives will also allow commercial use. This is the license used by Wikipedia, and is recommended for materials that would benefit from incorporating content from Wikipedia and similarly licensed projects.” Creative Commons, [CC BY-SA](#), 2020-09-06.
 - d. **CC BY-ND** - “This license lets others reuse the work for any purpose, including commercially; however, it cannot be shared with others in adapted form, and credit must be provided to you.” Creative Commons, [CC BY-ND](#), 2020-09-06.
 - e. **CC BY-NC** - “This license lets others remix, adapt, and build upon your work non-commercially, and although their new works must also acknowledge you and be non-commercial, they don’t have to license their derivative works on the same terms.” Creative Commons, [CC BY-NC](#), 2020-09-06.
 - f. **CC BY-NC-SA** - “This license lets others remix, adapt, and build upon your work non-commercially, as long as they credit you and license their new creations under the identical terms.” Creative Commons, [CC BY-NC-SA](#), 2020-09-06.
 - g. **CC BY-NC-ND** - “This license is the most restrictive of our six main licenses, only allowing others to download your works and share them with others as long as they credit you, but they can’t change them in any way or use them commercially.” Creative Commons, [CC BY-NC-ND](#), 2020-09-06.
6. **Open Data Commons** - “Open Data Commons is the home of a set of legal tools and licenses to help you publish, provide and use open data.” Open Data Commons, [Home](#), 2020-09-07.

Open Data Commons licensing include the following:

- a. **Open Data Commons Open Database License (ODbL)** - End users are free to *share*: To copy, distribute and use the database; *create*: To produce works from the

- database; and *adapt*: To modify, transform and build upon the database, as long as end users provide share-alike attribution and keeps open access for derivative works. Open Data Commons, [ODbL](#), 2020-09-07.
- b. Open Data Commons Attribution License (ODC-By) - End users are free to *share*: To copy, distribute and use the database; *create*: To produce works from the database; and *adapt*: To modify, transform and build upon the database, as long as end users provide attribution for derivative works. Open Data Commons, [ODC-By](#), 2020-09-07.
 - c. Open Data Commons Public Domain Dedication and License (PDDL) - End users are free to *share*: To copy, distribute and use the database; *create*: To produce works from the database; and *adapt*: To modify, transform and build upon the database with no restrictions. Open Data Commons, [PDDL](#), 2020-09-07.
7. **Unlicense** - “The Unlicense is a public domain equivalent license with a focus on an anti-copyright message. It was first published on January 1 (Public Domain Day), 2010. The Unlicense offers a public domain waiver text with a fall-back public-domain-like license, inspired by permissive licenses but without an attribution clause.” Wikipedia, [Unlicense](#), 2020-09-10.

Discussion

A software artifact written by an individual or a team within the United States of America, by definition, is covered under U.S. Copyright law under the “literary works” definition of [Title 17 U.S. Code § 101](#) (and by the [Bern Convention](#) more broadly), thus granting the author(s) exclusive rights to use and publish the software. In general, a copyright prevents others from legally using the software (albeit, this occurs all of the time). To transfer rights of use, software should be accompanied with a license that specifically dictates how it may be used by individuals or organizations other than the original author(s). There are many different types of software licenses that may be applied to software (see above for a general classification). The decision as to what software license to use depends on how the author(s) would like the principal work and any derivative work (i.e., modified works) to be treated by the end user. Details of each software license must be reviewed since subtle differences may or may not convey the intent of the author(s).

Interestingly, factual data (including observational data) and methodologies for collecting data are, in general, not protected by U.S. Copyright law. What may be covered under Copyright law with regard to data are creative ways to select or display data and the metadata that describe data. Although data themselves are not copyrightable, applying an explicit license to data for the purpose of sharing with the broader community can facilitate their reuse and further interpretation and understanding. Some licenses applied to data, however, may also complicate and inhibit sharing if the license is overly restrictive or requires excessive requirements for compliance.

To promote the fullest opportunity for sharing data and its inherent value, EDI recommends (and defaults to when the data package contributor does not include an explicit license) the [Creative Commons "no copyright reserved" \(CC0\) license](#) for data packages, which places the data package and its components into the public domain - in other words, no restrictions on use of the data package or its components. This does not imply that EDI condones use of any data package without attribution to its author(s). In contrast, EDI emphatically endorses the position that data packages (and their components) be properly cited to attribute the original work of author(s) and consider attribution both a community normative and a practice of professional etiquette. Under this arrangement, software artifacts that are a component of the data

package (as uploaded and published in the EDI data repository) will also fall under the CC0 license, thus adding the software to the public domain. As such, it is EDI's opinion that no other licensing strategy needs to be considered in this case. *[Note: other than CC0, CC licenses assume creative works are under copyright protections and therefore, should not be used for licensing data that cannot be copyrighted (see this Open Data Commons [FAQ - WHY NOT USE A CREATIVE COMMONS \(OR FREE/OPEN SOURCE SOFTWARE LICENSE\) FOR DATA\(BASES\)?](#)). For this reason, EDI may consider changing to the Open Data Commons [Public Domain Dedication and License](#) for data in the future.]*

If, however, a more restrictive license is desired by the author(s) of a software artifact that is included as a component of a data package, EDI recommends that such a licence should be explicitly stated (or referenced as being applied to the work) within the body of the software document, and preferably at the beginning of the document. Note that it is not clear whether this position is defensible from a legal perspective and may only be relevant to enforcing community best practices.

An alternative, and perhaps more legally defensible, approach would be to disassociate software artifacts from the data package and reference them solely through the data package metadata, thereby leaving them in a separate repository such as [GitHub](#). In this case, application of a software license should follow the recommendations of the host software repository. When in doubt, however, placing the license (or reference to the license) in the body of the software document is encouraged, as end users of the software artifact will visibly see the license upon inspection of the document.

Recommendation

As with data packages that are published through the EDI data repository, EDI recommends that all software related to contributed data be released into the public domain using either [CC0](#) or the [Unlicense](#) (although the Unlicense is now recognized by the Open Source Initiative as of June 2020, the Free Software Foundation believes [CC0 is more thorough and mature](#)). The rationale for this decision is primarily based on the fact that there exists a plethora of software licenses available to use (over 100 listed on the [GNU.org licensing website](#)), from public domain to permissive to copyleft to proprietary licensing, with each varying in their specification of which an end user would need to understand before employing the software in any derivative work. By releasing software into the public domain, the burden on the end user to fully understand the details of the software license is completely eliminated. As with any data package, EDI also believes it important to attribute the author(s) of the original work even when using public domain software. As noted above, any software artifact included as a component to a data package that is released into the public domain through CC0 will also be in the public domain. If, however, explicit licensing is required by the author(s) of software, EDI recommends the use of a “permissive” software license that allows for more creative and proprietary use of the software in derivative work, including its modification and embedding in closed source software (e.g., private or proprietary packages). In this case, EDI supports the recommendation of the [Free Software Foundation](#) (see [GNU.org licensing website](#)) to use the [Apache License, Version 2.0](#), a permissive and free software license that is compatible with version 3 of the GNU GPL.